

REPORT DOCUMENTATION PAGE			Form Approved OMB NO. 0704-0188		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY)		2. REPORT TYPE New Reprint		3. DATES COVERED (From - To) -	
4. TITLE AND SUBTITLE Sensitivity Analysis of Compressive Sensing Solutions		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER W911NF-04-D-0003			
		5c. PROGRAM ELEMENT NUMBER 611102			
6. AUTHORS Liyi Dai		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES North Carolina State University 2701 Sullivan Drive Admin Services III; Box 7514 Raleigh, NC 27695 -7514			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211			10. SPONSOR/MONITOR'S ACRONYM(S) ARO		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 62483-CS-SR.10		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT The compressive sensing framework finds a wide range of applications in signal processing and analysis. Within this framework, various methods have been proposed to find a sparse solution $x$ from a linear measurement model $y = Ax$ . In practice, the linear model is often an approximation. One basic issue is the robustness of the solution in the presence of uncertainties. In this paper, we are interested in compressive sensing solutions under a general form of measurement $y = (A + B)x + v$ in which $B$ and $v$ describe modeling and measurement inaccuracies, respectively. We analyze the sensitivity of solutions to infinitesimal modeling error $B$ or measurement inaccuracy $v$ . Exact					
15. SUBJECT TERMS compressive sensing, sparse solutions, sensitivity analysis, robustness					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Liyi Dai
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 919-549-4350

## Report Title

### Sensitivity Analysis of Compressive Sensing Solutions

#### ABSTRACT

The compressive sensing framework finds a wide range of applications in signal processing and analysis. Within this framework, various methods have been proposed to find a sparse solution  $x$  from a linear measurement model  $y = Ax$ . In practice, the linear model is often an approximation. One basic issue is the robustness of the solution in the presence of uncertainties. In this paper, we are interested in compressive sensing solutions under a general form of measurement  $y = (A + B)x + v$  in which  $B$  and  $v$  describe modeling and measurement inaccuracies, respectively. We analyze the sensitivity of solutions to infinitesimal modeling error  $B$  or measurement inaccuracy  $v$ . Exact solutions are obtained. Specifically, the existence of sensitivity is established and the equation governing the sensitivity is obtained. Worst-case sensitivity bounds are derived. The bounds indicate that sensitivity is linear to measurement inaccuracy due to the linearity of the measurement model, and roughly proportional to the solution for modeling error. An approach to sensitivity reduction is subsequently proposed.

---

## REPORT DOCUMENTATION PAGE (SF298) (Continuation Sheet)

---

Continuation for Block 13

ARO Report Number 62483.10-CS-SR  
Sensitivity Analysis of Compressive Sensing So...

Block 13: Supplementary Note

© 2015 . Published in Frontiers in Robotics and AI: Sensor Fusion and Machine Perception, Vol. Ed. 0 2, (1) (2015), (, (1). DoD Components reserve a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so (DODGARS §32.36). The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

Approved for public release; distribution is unlimited.

# Sensitivity Analysis in RIPless Compressed Sensing

Abdolreza Abdolhosseini Moghadam, Mohammad Aghagolzadeh, Hayder Radha

**Abstract**—Sensitivity analysis in optimization theory explores how the solution to a particular optimization problem changes as the objective function or constraints of such optimization problem perturb. A recent and yet important class of optimization problems is the framework of compressed sensing where the objective is to find the sparsest solution to an under-determined and possibly noisy system of linear equations. In this paper, we show that by utilizing some tools in sensitivity analysis, namely Invariant Support Sets (ISS), one can improve certain developed results in the field of compressed sensing. More specifically, we show that in a noiseless and RIP-less setting [11], the recovery process of a compressed sensing framework is a binary event in the sense that either all vectors with the same support and sign pattern can be recovered from their compressive samples or none can be estimated correctly. However, in a noisy and RIP-less setting, recovering only one signal from its limited noisy samples guarantees that there exist signals (possibly even with different supports and sign patterns) and noise vectors that shall be recovered with good accuracies by using Lasso.

**Index Terms**—sensitivity analysis, compressed sensing, compressive sampling, Invariant Support Set.

## I. INTRODUCTION

An important optimization problem which appears in a wide range of applications [1], [2], [3], [4], [5] is the problem of compressed sensing [6], [7], [8], [9], [10], [11], where the objective is to recover the sparsest (i.e. mostly zero) vector  $x \in \mathbb{R}^n$  from limited linear samples  $y = Ax$  where  $A \in \mathbb{R}^{m \times n}$ , and by limited we mean that the number of samples is less than the size of the vector  $x$ , i.e.  $m < n$ . On the other hand, sensitivity analysis [13], [14], [16] in optimization problems and systems of equations in general, investigates how the solution to an optimization problem or that system of equation changes by perturbation in constraints or the objective function. Despite the importance of such notion in the post calculation and analytical phase, to the best of our knowledge, no attempts has been made so far to deploy that kind of analysis in the compressed sensing framework. In this paper, we show that certain sensitivity analysis tools improve some solid results already developed in the field of compressed sensing.

It is well known that the success of finding the sparsest solution that satisfies the system of linear equation  $y = Ax$  heavily depends on the quality of the sensing matrix  $A$ . Early papers in the field of compressed sensing [6], [7] showed that if the matrix  $A$  has a property called Restricted Isometry (RIP), then the sparsest solution to the noiseless system of

$y = Ax$  has the minimum  $\ell_1$  norm, and hence the sparsest vector could be found by linear programming. In a noisy setting, where equations are contaminated by the noise vector  $\epsilon$  (i.e.  $y = Ax + \epsilon$ ), a popular technique (which provably works) to recover  $x$  from  $y$  is unconstrained Lasso [12], which can be cast as a quadratic programming optimization problem.

After the introduction of RIP, other sufficient conditions to guarantee successful recovery were proposed for the matrix  $A$  (for instance Null Space Property (NSP) [17]). However, proving or verifying that a matrix has RIP or NSP with the minimal number of rows ( $m$ ) turns out to be quite challenging or even impossible except for a few ensembles of well-known matrices such as matrices with entries following a Gaussian distribution. In the more recent work of [11], authors showed that as long as the sensing matrix  $A$  has isotropy and incoherence properties (see [11] for definitions), then the matrix  $A$  should be a “good” sensing matrix candidate for the compressed sensing framework. What signifies that work is the fact that, those conditions are much weaker than RIP or NSP and more importantly much easier to verify. Clearly achieving such strong result in [11] should not be free and indeed, the universality of matrices with RIP is given up in that work<sup>1</sup>. More specifically, that work shows that the probability of recovering a fixed but arbitrary signal  $x$  from compressive samples  $y = Ax$  is high (for instance at least  $1 - O(1)/n$  where  $n$  is the length of the unknown vector  $x$ ), but such matrix may not recover all imaginable sparse vectors. In this paper, by using simple arguments, we show that isotropy and incoherence properties have even greater consequences and in fact, recovering only a fixed but arbitrary signal from its compressive samples deterministically guarantees that there are possibly infinitely many other signals, sometimes even with different supports (from the recovered one) which could be recovered from their compressive samples as well and still the quality of recovery will be high. To that end, we use the notion of “Invariant Support Set” (ISS) [13], [14]. Broadly speaking, ISS studies under which perturbations in the constraints of an optimization problem, the support of the solution to that perturbed optimization problem stays unchanged. In a noiseless scenario, we show that, if the compressed sensing framework can exactly recover a signal  $x$  from pure compressive samples  $y = Ax$  then any other signal  $u$  with the same support and sign pattern of  $x$ , could be recovered exactly from  $q = Au$ . This extends the applicability of

\*Abdolreza Abdolhosseini Moghadam (abdolhos@msu.edu, corresponding author), Mohammad Aghagolzadeh (aghagol1@msu.edu) and Hayder Radha (radha@msu.edu) are with the Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA.

<sup>1</sup>Universality of a pair of sensing matrix and a decoder means that, the pair of decoder and the fixed sensing matrix could recover any sufficiently sparse vector from its compressive samples.

[11] from “a fixed but arbitrary signal” to “all signals with the same support and sign pattern”. In the noisy scenario, we show that if the compressed sensing framework has successfully recovered a signal  $x$  from noisy compressive samples  $y = Ax + \epsilon$  and if such recovery was not by chance<sup>2</sup>, then there exist infinitely many other signals (even with supports different from  $x$ ) and noise vectors, for instance  $u$  and  $\xi$  respectively for which, Lasso [12] returns an accurate estimate of  $u$  from noisy samples  $q = Au + \xi$ .

This paper is organized as follows: in Section II, we show how ISS might be utilized for the sensitivity analysis of the compressed sensing framework. Section III presents some numerical simulations and Section IV concludes the paper.

## II. SENSITIVITY ANALYSIS IN COMPRESSED SENSING

Before proceeding to the main results, we introduce some notations used in this paper. For a natural number  $m$ , we define the set  $[m] := \{1, 2, \dots, m\}$ . For sub-matrix selection, we adopt Matlab notations. For instance, let  $A \in \mathbb{R}^{m \times n}$  be an  $m \times n$  real-valued matrix,  $p \subseteq [m]$  and  $q \subseteq [n]$ . Then by  $A_{:,p}$  (and  $A_{q,:}$  respectively) we mean sub-matrix of  $A$  which is formed by selecting columns (and rows) indexed by those in  $p$  (and  $q$ ). Cardinality of a set  $l$  is denoted by  $\text{Card}\{l\}$ . The support of a vector  $t$ , denoted by  $\text{Support}\{t\}$  is the set of indices where  $t$  is non-zero, i.e.  $\text{Support}\{t\} := \{i : t_i \neq 0\}$ . By  $l^c$  we mean the complement of the set  $l$ . For simplicity of notation and without loss of generality, in this paper we have assumed that the unknown vector of interest to be recovered is non-negative ( $x_i, u_i \geq 0, \forall i \in [n]$ ). The extension to the general case, where signals can take negative values is straightforward.

We build our contribution on top of the RIPless theory of [11] where we assume that entries of the sensing matrix  $A \in \mathbb{R}^{m \times n}$  are i.i.d from a distribution  $F$  with the following two conditions:

- 1) Isotropy condition: The distribution  $F$  is zero-mean with variance  $1/m$ .
- 2) Incoherence property: If  $a \sim F$  then either deterministically or with high probability  $|a|^2 < \mu(F)$  where  $\mu(F)$  is the “coherence parameter” of the distribution of  $F$ . The smaller this value is, the fewer number of samples ( $m$ ) would be required for the recovery process. As in [11], we assume that the matrix  $A$  has  $\mu(F) = O(1/\sqrt{m})$ .

In the following two subsections, we discuss consequences of sensitivity analysis and existence of invariant support sets in the noiseless and noisy recovery cases, respectively.

### A. Noiseless setting

Assume that,  $A \in \mathbb{R}^{m \times n}$  and  $x$  is a non-negative  $k$ -sparse vector ( $\|x\|_0 = k$ ) of length  $n$  where  $k < m < n$ . In the noiseless case, a compressed sensing framework estimates the original vector  $x$  from samples  $y = Ax$  as  $\hat{x}$  where

<sup>2</sup>In the sense that,  $A$  was ill-conditioned but  $\|Ax\| \approx \|x\|$  for that particular  $x$ .

$$\hat{x} = \arg \min_z \mathbf{1}^T z \text{ s.t. } y = Az, z \succeq 0 \quad (1)$$

Here,  $\mathbf{1}^T$  is an all-one row vector with size  $n$  and  $z \succeq 0$  implies that all entries of  $z$  are non-negative.

Let  $LP(A, y)$  be the solution to the linear programming optimization problem of (1), i.e.  $LP(A, y)$  is a vector with the minimal  $\ell_1$  norm that satisfies  $y = A\hat{x}$ . The ISS interval of  $LP(A, y)$  is the set of all vectors  $\{\Delta y \in \mathbb{R}^m\}$  such that  $\text{Support}\{LP(A, y)\} = \text{Support}\{LP(A, y + \Delta y)\}$ .

The following Theorem proves that given a fixed sensing matrix  $A$  and in a noiseless setting, the recovery process of all vectors with the same support and sign pattern is a binary event in the sense that either all or none of signals with the same sign pattern and support could be recovered from their compressive samples. In other words, the recovery process in the noiseless setting is independent of the values of the signal, and instead depends on the support and sign pattern of that signal which sounds reasonable.

**Theorem 1.** *Assume that the matrix  $A$  has weak-RIP property in [11] (or estimates E1 – E4 in the same paper) and there exist a vector  $x$  with  $\text{Support}\{x\} = T$  such that  $LP(A, y = Ax) = x$ . Then, for any other vector  $u$  with the same support  $\text{Support}\{u\} = T$  and sign pattern, we would have:*

$$LP(A, q = Au) = u$$

*Proof.* Let  $\hat{x} = LP(A, y)$ . Then by assumptions of the theorem,  $\hat{x} = x$  and of course  $\text{Support}\{\hat{x}\} = \text{Support}\{x\} = T$ . If the perturbation vector of  $\Delta y = Au - Ax$  belongs to the ISS interval of  $LP(A, y)$  then it means that if we solve the perturbed problem of  $\hat{u} = LP(A, y + \Delta y) = LP(A, q)$  then  $\text{Support}\{\hat{u}\} = \text{Support}\{\hat{x}\} = T$  by the definition of ISS. In words, in this case, the support of the vector is estimated exactly correct (note that  $\text{Support}\{u\} = T$  as well). If that happens, since  $A$  is non-singular on column indices  $T$  (in fact, assumptions of E1 – E4 in [11] imply that  $A_{:,p}$  is near isometry in the noiseless case), thus  $\hat{u}$  must be  $u$ . To that end, by theorem 2.5 in [13] we need to find a non-negative vector  $r$  such that  $A_{:,T} r_T = y + \Delta y = Au$ . Clearly that vector exist which is simply  $r = u$ , since by the assumption  $\text{Support}\{u\} = T$  and  $u \succeq 0$ . Now, in a general case when signals can take negative values, the compressed sensing formulation for the recovery process is

$$\hat{x} = \arg \min_z \|z\|_1 \text{ s.t. } y = Az$$

However, one can re-format the above equation into the canonical form of

$$\begin{aligned} \hat{x}_p, \hat{x}_n &= \arg \min_{z_p, z_n} \mathbf{1}^T \begin{bmatrix} z_p \\ z_n \end{bmatrix} \\ \text{s.t. } y &= \begin{bmatrix} A & -A \end{bmatrix} \begin{bmatrix} z_p \\ z_n \end{bmatrix}, \quad z_p, z_n \succeq 0 \end{aligned}$$

and form the final estimate as  $\hat{x} = \hat{x}_p - \hat{x}_n$ . It should be clear that in this general case, we are still working with non-negative vectors  $\hat{x}_p$  and  $\hat{x}_n$  and consequently, our ISS

arguments for the formulation of (1) will hold in this case as well. ■

Note that Theorem 1 is more general than the notion of “a fixed but arbitrary signal” in [11] which derives the probability of recovery for only one fixed signal  $x$ . Although, it can be argued that all signals with the same support and sign pattern share the same exact and inexact dual certificates (see [11] for definitions) and using that fact, one can derive a bound on the error of the estimate  $\|\hat{u} - u\|$ . Nevertheless and as shown above, our proof is much simpler due to the sensitivity analysis.

In the following, we prove that if the recovery of  $x$  from noisy samples  $y = Ax + \epsilon$  had been successful (in the sense that the estimate  $\hat{x}$  has the property that  $\|x - \hat{x}\| = O(\epsilon)$ ), then either deterministically or at least with much higher probabilities than [11], infinitely many other signals (which possibly could have supports different from  $x$ ) could be accurately recovered from their limited compressive samples as well.

### B. Noisy setting

For the noisy case and also when the vectors to be recovered are compressible rather than exactly sparse, an approach to recover a signal  $x$  from noisy measurements  $y = Ax + \epsilon$  is utilizing the unconstrained Lasso [12] formulation:

$$\hat{x} = \arg \min_b \frac{1}{2} \|y - Ab\|_2^2 + \lambda \|b\|_1 \quad (2)$$

Here,  $\lambda$  is a fixed regularization parameter that can balance the sparsity of the estimation  $\|\hat{x}\|_0$  against the discrepancy of the estimate  $y - A\hat{x}$ . Again let us consider the case when the unknown vector  $x$  is assumed to be non-negative and exactly sparse. Let  $QP(A, y, \lambda)$  be the solution to the following equivalent but canonical unconstrained Quadratic Programming (QP) problem of (2):

$$\hat{x} = \arg \min_b C^T b + \frac{1}{2} b^T Q b \quad \text{s.t. } b \succeq 0 \quad (3)$$

Where

$$C = -A^T y + \lambda \mathbf{1}, Q = A^T A \quad (4)$$

The dual problem of above is in the form of

$$\begin{aligned} \hat{x}, \hat{s} &= \arg \max_{b, s} -\frac{1}{2} b^T Q b \\ \text{s.t. } s - Qb &= C, \quad b, s \succeq 0 \end{aligned} \quad (5)$$

In summary, to recover  $x$  from noisy samples  $y = Ax + \epsilon$ , one declares  $\hat{x} = QP(A, y, \lambda)$  as an estimate of  $x$  given a fixed regularization parameter  $\lambda$ . Let us use  $p$  to denote the support of the estimate  $\hat{x}$  ( $p = \text{Support}\{\hat{x}\} = \text{Support}\{QP(A, y, \lambda)\}$ ). Note that, since this time we are working in the noisy realm, the event that the support of the estimate is exact (i.e. the event of  $\text{Support}\{x\} = p$ ) is very unlikely and in practice  $p$  is much larger than the support of the sparse vector  $x$ . In fact, in the noisy setting,  $p$  is not necessarily a subset of  $\text{Support}\{x\}$ . However, the converse ( $\text{Support}\{x\} \subset p$ ) usually holds if the recovery process has

been successful. Here, we assume that this case happens in the recovery of  $x$  (see [19], [20] for some conditions on the support recovery). Now, let us find ISS interval of  $QP(A, y, \lambda)$  which is the set:

$$\left\{ \Delta y : \text{Support}\{QP(A, y, \lambda)\} = \text{Support}\{QP(A, y + \Delta y, \lambda)\} \right\}$$

Note that, this set is always non-empty as the trivial zero solution for  $\Delta y$  belongs to this set. Clearly, we are interested in the case when  $\Delta y$  is non-zero. We show that this would be the case and more importantly, if the recovery of  $x$  had been successful in the sense that

$$\|x - QP(A, y, \lambda)\| = O(\|\epsilon\|)$$

then, corresponding problems in the ISS interval of  $QP(A, y, \lambda)$  could be solved with good precision as well in the same sense. In contrast to the noiseless setting and as stated before, the recovery of a vector  $x$  from noisy samples  $y = Ax + \epsilon$  could guarantee the recovery of signals with supports possibly different from  $\text{Support}\{x\}$ . The benefit of using such sensitivity analysis is that otherwise for each new support and sign pattern one must find exact or inexact dual certificates [11] which exist with a probability of (at least)  $1 - O(1)/n$ . Although such probability is high, it might not be sufficiently high or comparable with our result. This can be shown by noting that  $\text{Card}\{\text{Support}\{\hat{x}\}\}$  could be as high as  $m$  and there are  $\binom{m}{k}$  possible cases just for supports of size  $k$  within  $\text{Support}\{\hat{x}\}$ , and hence, the probability that dual certificates for all these possible supports exist could be very small at the end (if the probability of  $1 - O(1)/n$  is used without any improvement). In contrast, steps in our approach mostly hold either deterministically or at least with much higher probabilities. And finally, our error bound could have smaller constants when compared to the original work in [11] when proper regularization parameters are set and the matrix  $A$  is well conditioned.

To utilize results from Invariant Support Sets, we need the following lemma which states  $Q_{pp} := A_{:,p}^T A_{:,p}$  has an inverse for  $p = \text{Support}\{QP(A, y, \lambda)\}$ .

**Lemma 2.** *Let  $\hat{x}$  be the unique minimizer to  $QP(A, y, \lambda)$  and define  $p = \text{Support}\{\hat{x}\}$ . Then  $\text{Card}\{p\} \leq m$  and  $A_{:,p}$  is full rank.*

*Proof.* For now assume that  $\text{Card}\{p\} > m$  or  $A_{:,p}$  is not full-rank. In both cases, the null-space of  $A_{:,p}$  is not empty and there exists a vector  $h \subseteq p$  such that  $A_{:,p} h = 0$ . Define

$$t := \min_{i \in \text{Support}\{h\}} \left| \frac{\hat{x}_i}{h_i} \right|$$

Now consider two vectors of  $\hat{x} + th$  and  $\hat{x} - th$ . Let  $l = \{i : h_i > 0\}$  be indices where  $h$  is positive and define  $H := t|h|$  (i.e.  $H_i = th_i$  for  $i \in l$  and  $H_i = -th_i$  when  $i \in \text{Support}\{h\}/l$ ). Note that all entries of  $\hat{x}_p$  are always non-negative and  $\hat{x}_i \geq H_i$ . Therefore,  $|\hat{x}_i + h_i| = \hat{x}_i + H_i$  when  $i \in l$  and  $|\hat{x}_i + h_i| = \hat{x}_i - H_i$  in  $i \in l^c = \text{Support}\{h\}/l$ .

Since  $\hat{x}$  was the unique solution to the optimization problem of  $QP(A, y, \lambda)$ , we have

$$\frac{1}{2}\|y - A\hat{x}\|_2^2 + \lambda\|\hat{x}\|_1 < \frac{1}{2}\|y - A(\hat{x} + th)\|_2^2 + \lambda\|\hat{x} + th\|_1$$

and also

$$\frac{1}{2}\|y - A\hat{x}\|_2^2 + \lambda\|\hat{x}\|_1 < \frac{1}{2}\|y - A(\hat{x} - th)\|_2^2 + \lambda\|\hat{x} - th\|_1$$

However, since  $h$  is in the null-space of  $A$ , we have  $A(\hat{x} + th) = A(\hat{x} - th) = A\hat{x}$ . This means that  $\|\hat{x}\|_1 < \|\hat{x} + th\|_1$  and  $\|\hat{x}\|_1 < \|\hat{x} - th\|_1$  simultaneously. Assuming  $\|\hat{x}\|_1 < \|\hat{x} + th\|_1$  means

$$\sum_{i \in p} \hat{x}_i < \sum_{i \in l} (\hat{x}_i + H_i) + \sum_{i \in l^c} (\hat{x}_i - H_i)$$

and hence  $\|H_l\|_1 > \|H_{l^c}\|_1$ . Similarly  $\|\hat{x}\|_1 < \|\hat{x} - th\|_1$  means that

$$\sum_{i \in p} \hat{x}_i < \sum_{i \in l^c} (\hat{x}_i + H_i) + \sum_{i \in l} (\hat{x}_i - H_i)$$

This translates to  $\|H_{l^c}\|_1 > \|H_l\|_1$  which contradicts  $\|\hat{x}\|_1 < \|\hat{x} + th\|_1$ . This means that, if  $\text{Card}\{p\} > m$  or  $A_{:,p}$  is not full-rank then one can find a vector  $z$  (in here either  $th$  or  $-th$ ) in the null-space of  $A_{:,p}$  such that  $\|\hat{x} + z\|_1 < \|\hat{x}\|_1$ , attaining a lower objective function of  $\frac{1}{2}\|y - A(\hat{x} + z)\|_2^2 + \lambda\|\hat{x} + z\|_1$  which contradicts the assumption that  $\hat{x}$  is the unique minimizer to  $QP(A, y, \lambda)$ . Thus both  $\text{Card}\{p\} \leq m$  and  $A_{:,p}$  is full-rank. ■

In the following theorem, we derive characteristics of signals  $u$  and noise vectors  $\xi$  that lead to perturbations  $\Delta y = A(u - x) + (\xi - \epsilon)$  residing in the ISS of  $QP(A, y = Ax + \epsilon, \lambda)$ .

**Theorem 3.** Assume that the matrix  $A$  has weak-RIP property in [11] (or estimates E1 – E4 in the same paper). Furthermore, assume that  $\hat{x}$ , the unique minimizer to  $QP(A, y = Ax + \epsilon, \lambda)$  satisfies  $\|x - \hat{x}\| = O(\|\epsilon\|)$ . Then ISS of  $QP(A, y, \lambda)$  is not empty if  $\|A_{:,p}\|$  is sufficiently small for  $p := \text{Support}\{\hat{x}\}$ .

*Proof.* By Lasso formulation and our assumptions  $\forall i \in p : \hat{x}_i > 0$ . We need to find for which signal  $u$  and noise vector  $\xi$ ,  $\text{Support}\{\hat{u} := QP(A, q = Au + \xi, \lambda)\}$  would stay at  $p$ . To that end, let us define:

$$\begin{aligned} C^{\{x\}} &= -A^T y + \lambda \mathbf{1} \\ C^{\{u\}} &= -A^T q + \lambda \mathbf{1} \\ \Delta x &= u - x \\ \Delta \epsilon &= \xi - x \\ \Delta C &= C^{\{u\}} - C^{\{x\}} = A^T(y - q) = -A^T(A\Delta x + \Delta \epsilon) \\ z &= \{1, 2, \dots, n\}/p \\ Q_{pz} &= A_{:,p}^T A_{:,z} \\ Q_{pp} &= A_{:,p}^T A_{:,p} \end{aligned}$$

In summary, if  $\Delta y := q - y = A\Delta x + \Delta \epsilon$  belongs to the ISS of  $QP(A, y, \lambda)$ , then  $\text{Support}\{QP(A, q = y + \Delta y, \lambda)\} = p$  by definition. As shown in [14], to find such ISS interval, it is required to find non-negative vectors  $\hat{u}$  and  $s^{\{u\}}$  which satisfy the following two equations

$$\begin{aligned} -Q_{pp}\hat{u}_p &= C_p^{\{u\}} = C_p^{\{x\}} + \Delta C_p \\ s_z^{\{u\}} &= Q_{pz}^T \hat{u}_p + C_z^{\{u\}} \end{aligned} \quad (6)$$

Note that, since  $\hat{x} = QP(A, y, \lambda)$ , there exists a non-negative vector  $s^{\{x\}}$ , with the support on  $z = [n]/p = p^c$  such that

$$-Q_{pp}\hat{x}_p = C_p^{\{x\}}, \quad s_z^{\{x\}} = Q_{pz}^T \hat{x}_p + C_z^{\{x\}}$$

Since  $A_{:,p}$  is full rank (see Lemma 2),  $Q_{pp} = A_{:,p}^T A_{:,p}$  has an inverse. Consider the vector  $\hat{u}$  which at indices  $p$  equals to

$$\hat{u}_p = (-Q_{pp})^{-1} C_p^{\{u\}}$$

and is zero in other indices (i.e.  $z = [n]/p$ ). As [16], let us use  $v^{\{i\}}$  to denote the  $i$ -th row of  $-Q_{pp}^{-1}$  and note that

$$\forall i \in p : \hat{x}_i = v^{\{i\}} C_p^{\{x\}} > 0$$

Similarly, for all  $i \in p$

$$\hat{u}_i = v^{\{i\}} C_p^{\{u\}} = v^{\{i\}} (C_p^{\{x\}} + \Delta C_p) = \hat{x}_i + v^{\{i\}} \Delta C_p \quad (7)$$

Using full-rank property of  $Q_{pp} = A_{:,p}^T A_{:,p}$ , one can represent any vector in  $\mathbb{R}^{\text{Card}\{p\}}$  in terms of columns of  $Q_{pp}$ . Let us define

$$\Delta \hat{\epsilon} := -Q_{pp}^{-1} A_{:,p}^T \Delta \epsilon = -A_{:,p}^+ \Delta \epsilon, \quad \Delta \hat{x} := -Q_{pp}^{-1} \Delta x$$

where by  $A_{:,p}^+$  we mean the pseudo-inverse of  $A_{:,p}$ . Clearly  $\|\Delta \hat{\epsilon}\| \leq \|A_{:,p}^+\| \|\Delta \epsilon\|$  and  $\|\Delta \hat{x}\| \leq \|Q_{pp}^{-1}\| \|\Delta x\|$ . Since the recovery of  $x$  has been successful and the support of the estimate  $\hat{x}$  was  $p$ , hence it is very unlikely that  $\|A_{:,p}\|$  (and hence  $\|Q_{pp}^{-1}\|$  and  $\|A_{:,p}^+\|$ ) to be very high. In fact, if  $\text{Card}\{p\} - k$  is not too big, then  $A_{:,p}$  has good condition number due to weak-RIP. Nevertheless, plugging these values into (7) yields:

$$\hat{u}_i = \hat{x}_i + \Delta \hat{x}_i + \Delta \hat{\epsilon}_i \quad (8)$$

For now let us pause in here and focus on  $s_z^{\{u\}}$ , which is by definition:

$$\begin{aligned} s_z^{\{u\}} &= Q_{pz}^T \hat{u}_p + C_z^{\{u\}} \\ &= Q_{pz}^T (\hat{x} - Q_{pp}^{-1} \Delta C_p) + C_z^{\{x\}} + \Delta C_z \\ &= s_z^{\{x\}} - Q_{pz}^T Q_{pp}^{-1} \Delta C_p + \Delta C_z \end{aligned}$$

Using the same arguments for  $\Delta \hat{x}$ , one can define  $\Delta \hat{C} := Q_{pp}^{-1} \Delta C$  where clearly  $\|\Delta \hat{C}\| \leq \|Q_{pp}^{-1}\| \|\Delta C\|$ . Therefore

$$s_z^{\{u\}} \succeq s_z^{\{x\}} - (\|\Delta C_z\|_\infty + \|Q_{pz}^T \Delta \hat{C}_p\|_\infty) \mathbf{1}$$

Where by  $\succeq$ , we mean entry-wise greater than operator. By off support incoherency (see Lemma 2.4 in [11]), it is straightforward to verify that:

$$\Pr \left( \|Q_{pz}^T \Delta \hat{C}_p\|_\infty \geq t \|\Delta \hat{C}_p\|_2 \right) < n^{\frac{-3}{2 + \frac{6}{\mu \log n}}}$$

for

$$t = \frac{\mu \log n}{\sqrt{m}} = O\left(\frac{1}{k}\right)$$

As in [11], here we have assumed  $\mu$  (the coherency parameter of distribution of the entries of  $A$ ) is  $O(1/\sqrt{m})$ . Let us assume that  $\|Q_{pp}^{-1}\| \leq \frac{1}{t} = O(k)$ . Then by properties of the norm operator

$$\begin{aligned} s_z^{\{u\}} &\succeq s_z^{\{x\}} - (\|\Delta C_z\|_2 + t \|\Delta \hat{C}_p\|_2) \mathbf{1} \\ &\succeq s_z^{\{x\}} - (\|\Delta C_z\|_2 + \|\Delta C_p\|_2) \mathbf{1} \end{aligned} \quad (9)$$

Note that  $\|\Delta C_z\|^2 + \|\Delta C_p\|^2 = \|\Delta C\|^2$  and hence  $\|\Delta C_z\| + \|\Delta C_p\| \leq \sqrt{2} \|\Delta C\|$ , which gives us the final inequality

$$s_z^{\{u\}} \succeq s_z^{\{x\}} - \sqrt{2} \|\Delta C\|_2 \mathbf{1} \quad (10)$$

Having (10) and (8) and recalling that  $s_z^{\{x\}}$  and  $\hat{x}_p$  are both strictly positive, it should be clear that if  $\|\Delta C\|$  is not too big then one can find non-zero configurations for  $\Delta C$  (or equivalently signal  $u$  and noise vector  $\xi$ ) where  $s_z^{\{u\}}$  and  $\hat{u}_i$  will both be always positive which finally proves that ISS of  $QP(A, y, \lambda)$  is not empty. ■

In the following, we prove that with a proper regularization parameter  $\lambda$ , the corresponding signals in the ISS of an already recovered signal are guaranteed to be recovered with low errors as well.

**Theorem 4.** Suppose that  $\hat{x} = QP(A, y = Ax + \epsilon, \lambda)$  has the support on  $p = \text{Support}\{\hat{x}\}$  and  $\text{Support}\{x\} \subseteq p$  and  $\|x - \hat{x}\| = O(\|\epsilon\|)$ . Let  $q = Au + \xi$  and assume  $\text{Support}\{u\} \subseteq p$  and  $\Delta y = q - y$  belongs to ISS of  $QP(A, y, \lambda)$ . Then the error of the estimate of  $\hat{u} = QP(A, u, \lambda)$  is in the form of

$$\|u - \hat{u}\| \leq \|(A_{:,p}^T A_{:,p})^{-1}\| (\lambda \sqrt{m} + \|A_{:,p}^T\| \|\xi\|)$$

*Proof.* If  $\Delta y$  belongs to ISS of  $QP(A, y, \lambda)$ , then by definition of ISS,  $\text{Support}\{\hat{u} = QP(A, q, \lambda)\} = p$  as well and therefore

$$\hat{u}_p = -(A_{:,p}^T A_{:,p})^{-1} (-A_{:,p}^T q + \lambda \mathbf{1})$$

where  $\mathbf{1}$  is an all one vector with size of  $\text{Card}\{p\}$ . Recalling that  $q = Au + \xi = A_{:,p} u_p + \xi$  and the fact that  $A_{:,p}$  is full rank, it can be concluded that

$$\hat{u}_p = u_p + (A_{:,p}^T A_{:,p})^{-1} (A_{:,p}^T \xi - \lambda \mathbf{1})$$

Since  $u$  is supported on  $p$

$$\|\hat{u} - u\| \leq \|A_{:,p}^+\| \|\xi\| + \lambda \sqrt{m} \|(A_{:,p}^T A_{:,p})^{-1}\|$$

It should be noted that the term  $\lambda \sqrt{m}$  in the error bound (as opposed to what it seems), is not very large. Let us show that with an example. Without any loss of generality assume that we are recovering a unit-norm vector and as in [11], the regularization parameter  $\lambda$  equals to  $10\sigma\sqrt{\log n}$  where  $\sigma$  is the standard deviation of the entries in the noise (here the vectors  $\epsilon$  and  $\xi$ ). It should be clear that since we are working in a normalized setting (i.e.  $x$  is unit norm and the expected value of the gram matrix of  $A$  is the identity matrix), then pure noiseless samples  $Ax$  and  $Au$  should have a norm around one as well. Since the dimension of noise (either  $\epsilon$  or  $\xi$ ) is  $m$ , we expect that for a successful recovery, the variance of the noise distribution ( $\sigma^2$ ) to be considerably smaller than  $1/m$ , since otherwise, on average the energy of the noise is either equal or larger than pure samples ( $E(\|\epsilon\|_2^2) = E(\|\xi\|_2^2) = m\sigma^2 \geq 1$ ). Therefore, in this assumed normalized case, we expect that  $\lambda$  to be much smaller than  $10\sqrt{(\log n)/m}$  and consequently  $\lambda \sqrt{m}$  should be much smaller than  $10\sqrt{\log n}$ . Also, it is important to note that  $\lambda$  needs to be smaller than  $\|A^T y\|_\infty$  since otherwise for  $\lambda \geq \|A^T y\|_\infty$ ,  $\hat{x}$  would be fixed at zero which is clearly wrong [15].

**Corollary 5.** If  $A$  has weak-RIP of order  $\text{Card}\{p\}$ , then estimate of  $\hat{u}$  has the following error for a small positive constant  $\delta$

$$\|\hat{u} - u\| \leq \frac{1}{1 - \delta} (\lambda \sqrt{m} + \sqrt{1 + \delta} \|\xi\|)$$

*Proof.* By the definition of weak-RIP:  $\sqrt{1 - \delta} \leq \|A_{:,p}\| \leq \sqrt{1 + \delta}$ . Plugging such value into Theorem 4 proves the claim. ■

### III. SIMULATION RESULTS

For our simulations, we could have adapted well-studied matrices with RIP (e.g. Gaussian ensembles) as sensing matrices  $A$  since they surely would have isotropy and incoherence properties as well. However, we have intentionally utilized SERP matrices (introduced in [21]) in the sensing mechanism. The reason for such selection is that, SERP matrices do not have RIP of a proper order with  $m = O(k \log n)$  rows. However, they admit isotropy and incoherency properties (although the desirable value of  $\mu = O(1/\sqrt{m})$  is not achievable with  $m = O(k \log n)$ ). Broadly speaking, SERP matrices are real-valued sparse matrices where each column is non-zero in  $d = O(\log n)$  random row indices and each non-zero entry is a zero mean Gaussian random variable with the variance of  $1/d$ . We have set  $d = 5 \log n$  for matrix generation in our simulations. Also for solving  $LP(A, y)$  and  $QP(A, y, \lambda)$ , we have used L1-LS algorithm [18].

In practice, even when one attempts to recover an exactly sparse vector  $x \in \mathbb{R}^n$  from compressive samples  $y \in \mathbb{R}^m$  (whether it's noisy or noiseless),  $LP(A, y)$  and  $QP(A, y, \lambda)$  virtually always return compressible<sup>3</sup> estimates  $\hat{x}$ . This is mainly due to finite hardware precision, insufficient number

<sup>3</sup>By compressible we mean that most of entries have small magnitudes instead of being exactly zero.



of iterations in optimization algorithms and similar related issues. Hence if one wants to simulate our claims about ISS then it will be always observed that solutions to the perturbed and original optimization problems would be both non-zero on all entries. Therefore, ISS would trivially hold true, since for instance

$$\text{Support}\{LP(A, y)\} = \text{Support}\{LP(A, y + \Delta y)\} = [n]$$

On the other hand, variations in the support of the solution to an optimization problem might not be abstracted into a single number due to its dimension. To address these concerns, we needed to define some notations and functions, presented in here:

- 1) For a natural number  $z$  and a vector  $\hat{u}$ ,  $T(z, \hat{u})$  is the support of top  $z$  entries of  $\hat{u}$  with the largest magnitudes. Hence if  $H_{[z]}(\cdot)$  denotes the hard-thresholding operator which keeps  $z$  largest coefficients of its argument then,  $T(z, \hat{u}) = \text{Support}\{H_{[z]}(\hat{u})\}$ . Clearly  $T(i, \hat{u}) \subseteq T(i+1, \hat{u})$ .
- 2) For a natural number  $z$  and two vectors  $\hat{x}$  and  $\hat{u}$ , define

$$\begin{aligned} f(z, \hat{u}, \hat{x}) &= \min b \\ \text{s.t. } T(z, \hat{u}) &\subseteq T(b, \hat{x}) \end{aligned} \quad (11)$$

In other words, all top  $z$  largest entries of  $\hat{u}$ , can be found in top  $f(z, \hat{u}, \hat{x})$  largest entries of  $\hat{x}$  and furthermore,  $f(z, \hat{u}, \hat{x})$  is the minimal value with such property.

It can be easily verified that if magnitudes of two vectors  $\hat{u}$  and  $\hat{x}$  have exactly the same pattern, then  $f(z, \hat{u}, \hat{x})$  should be equal to  $z$ . Note that, this condition is much stronger than just to say that top  $l\%$  coefficients of  $\hat{u}$  and  $\hat{x}$  have the same support for some value of  $l$ . For instance assume that  $\hat{x} = [3 \ 2 \ 1 \ 0]^T$ . Then top three largest coefficients of  $\hat{u} = [2 \ 1 \ 3 \ 0]^T$  and  $\hat{x}$  have the same support. However, stating that  $f(z, \hat{u}, \hat{x}) = z$ , implies that  $|\hat{u}_1| \geq |\hat{u}_2| \geq |\hat{u}_3|$  (i.e. the same magnitude pattern in  $\hat{x}$ ).

Recall that in this paper,  $m$  is the number of rows in  $A$ ,  $n$  denotes the length of the unknown vector  $x$  (and clearly the number of columns of  $A$ ),  $k$  is the number of non-zero entries of  $x$ , the noise variance is shown by  $\sigma^2$  and the regularization parameter of Lasso is  $\lambda$ . Given a fixed set of parameters, we run the following algorithm: we generate a random SERP matrix  $A$ , a unit norm, non-negative vector  $x$  with a random support and magnitudes and a Gaussian noise vector  $\epsilon$ . Depending on the noise variance, we estimate it by  $\hat{x} = LP(A, y)$  or  $\hat{x} = QP(A, y, \lambda)$  for  $y = Ax + \epsilon$ . To generate the perturbed signal  $u$ , we randomly change  $x$  in  $\alpha$  indices among support of  $x$  and make it non-zero in  $\beta$  extra random indices among  $T(2k, \hat{x})/\text{Support}\{x\}$ . Hence,  $\text{Support}\{x\} \subseteq \text{Support}\{u\}$  and  $\text{Card}\{\text{Support}\{u\}\} = \text{Card}\{\text{Support}\{x\}\} + \beta$ . After this step, we normalize  $u$  and make sure that  $u$  would remain non-negative as well. In our simulation, there will be no perturbations in noise ( $\Delta\epsilon = \xi - \epsilon = 0$ ) to limit enormous degrees of freedom in simulations. To have an idea on how similar or distinct are vectors  $x$  and  $u$  in our simulation, we report the Signal

to Noise Ratio (SNR) between  $x$  and  $u = x + \Delta x$  in the following. Let  $\hat{u}$  be the estimate of a compressed sensing framework for  $u$  from samples  $q = Au + \epsilon$ . If ISS holds in that framework, then  $f(z, \hat{u}, \hat{x}) = z$ .

We have simulated each of the following three configurations for one thousand times. No attempt has been made to optimize the regularization parameter  $\lambda$  which was fixed at  $\lambda = 0.002$ :

**SIM1:**  $m = 240$ ,  $k = 40$ ,  $n = 1000$ ,  $\alpha = k$ ,  $\beta = 0$  and  $\sigma = 0$ . In this configuration, there is no noise in samples. Also,  $x$  and  $u$  are two independent unit-norm signals with the same support. However, at each of 1000 runs, both the support and also the matrix  $A$  would change randomly. The average and median SNR between unit norm vectors of  $u$  and  $x$  had been only 3.11 and 3.05 dB respectively. Hence although both signals had the same support, they were totally different signals in simulations. The correlation coefficient between SNR of  $u$  to  $\hat{u}$  and SNR of  $x$  to  $\hat{x}$  was computed at 0.986. This means that, whenever  $x$  was recovered with smaller errors, then the estimate of  $u$  would have smaller errors as well. This meaningfully high correlation when SNR of  $x$  to  $u$  is that low, acknowledges our claim that the quality of the recovery process in the noiseless and RIPless case is independent of the signal value and only depends on the support of signals.

**SIM2:**  $m = 600$ ,  $k = 100$ ,  $n = 5000$ ,  $\alpha = \beta = 2$  and  $\sigma^2 = 0.001$ . The average and median SNR for sampling process in all 2000 recoveries<sup>4</sup> both were approximately 32.30 dB. The average and median SNR between  $x$  and the perturbed vector  $u$  were respectively 19.01 and 18.07 dB. The correlation coefficient between SNR of recoveries of  $u$  and  $x$  was 0.995.

**SIM3:**  $m = 1200$ ,  $k = 200$ ,  $n = 3000$ ,  $\alpha = 10$ ,  $\beta = 3$  and  $\sigma^2 = 0.001$ . The average and median SNR for sampling process in all 2000 recoveries both were approximately 29.38 dB. The average and the median SNR between  $x$  and the perturbed vector  $u$  were 13.42 and 13.18 dB respectively. Hence, perturbed vectors  $u$  are distinctly different from  $x$  in this scenario as well. The correlation coefficient between SNR of recoveries of  $u$  and  $x$  was 0.996.

Left column of Figure 1 shows how similar signals  $x$  and  $u$  were in each iteration. Qualities of recoveries are presented in the middle column of Figure 1. The right column of the same figure investigates whether ISS holds for these simulations by utilizing the function of  $f(z, \hat{u}, \hat{x})$ , introduced in equation (11). As stated before, ideally we expect that  $f(z, \hat{u}, \hat{x}) = z$ . However, it should be noted that, the function  $f(z, \hat{u}, \hat{x})$  is sensitive to small permutations in supports of  $\hat{u}$  and  $\hat{x}$ . For instance, assume that three largest coefficients of  $\hat{u}$  and  $\hat{x}$  are at indices of  $a, b, c$  and  $b, c, a$  respectively. Then although these two vectors have the same support, we would have  $f(1, \hat{u}, \hat{x}) = f(2, \hat{u}, \hat{x}) = f(3, \hat{u}, \hat{x}) = 3$  which is away from  $f(z, \hat{u}, \hat{x}) = z$ . This should explain some straight, horizontal lines in our plot. Also, since in SIM2 and

<sup>4</sup>In each of one thousand iterations, one set of samples is measured for  $u$  and another set of samples is measured for  $x$ .

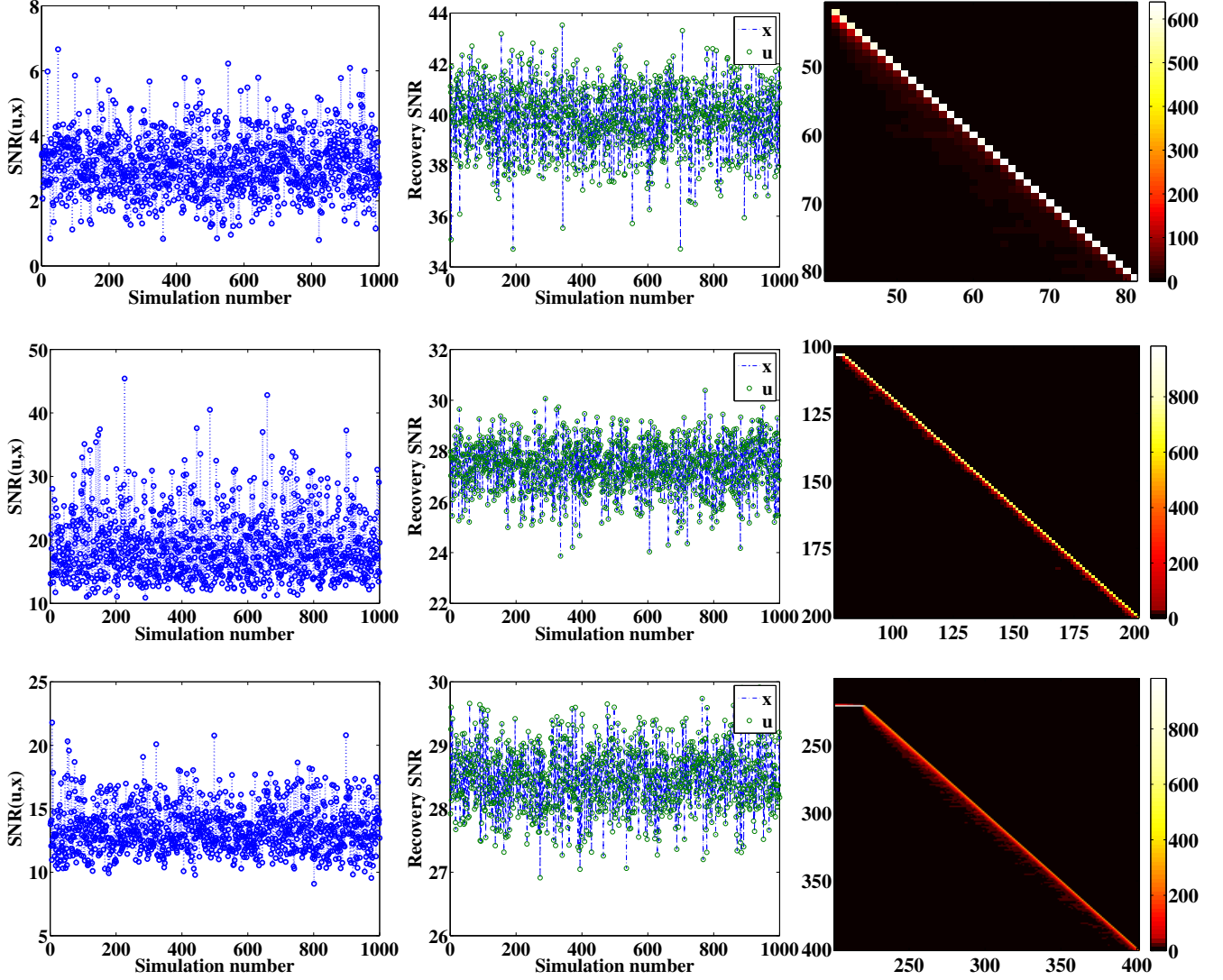


Fig. 1: Top row: SIM1, middle row: SIM2 and bottom row is SIM3. Left column: SNR (in dB) between the original signal  $x$  and the perturbed signal of  $u = x + \Delta x$  as a function of the simulation number. Middle column: SNR of recoveries for  $u$  and  $x$  as a function of the simulation number. Right column:  $f(z, \hat{u}, \hat{x})$  as a function of  $z$ . If the magnitude of the plot at row  $i$  and column  $j$  is  $t$ , it means that out of 1000 simulations, in  $t$  times  $f(j, \hat{u}, \hat{x}) = i$ .

SIM3,  $\text{Card}\{\text{Support}\{x\}\} < \text{Card}\{\text{Support}\{u\}\}$  then we expect that  $f(z, \hat{u}, \hat{x})$  to be fixed at  $\text{Card}\{\text{Support}\{u\}\}$  for small values of  $z$ , which explains the bright horizontal lines for those values of  $z$ . As a final note, our arguments in the previous section were for the worst case scenario and in practice there are many perturbed problems which are still in the ISS while they do not meet equations (8) and (10). On the other hand, we had generated perturbations of  $\Delta x = u - x$  randomly in our simulations. Consequently, true ISS conditions might be violated in our simulations, leading to anomalies in plots  $f(z, \hat{u}, \hat{x})$ . However, as it is clearly illustrated in Figure 1, the notion of ISS for compressed sensing holds to very good extent even in those random simulations since the function  $f(z, \hat{u}, \hat{x})$  looks like to be equal to  $z$  in most cases

#### IV. CONCLUSION

Invariant Support Set (ISS) of an optimization problem is the set of perturbations in the objective function or constraints of that problem, such that the support of the solution stays intact. In this paper, we showed that in a RIPless scenario for the compressed sensing framework, any problem which can be solved correctly, would have a non-empty ISS. This means that in the noiseless case, successful recovery of a fixed but arbitrary signal guarantees that all other signals with the same support and sign pattern could be recovered exactly as well. In a noisy setting however, good recovery of a fixed but arbitrary signal guarantees that there exist infinitely many other signals, (possibly even with different supports), which can be recovered with good precision from their noisy compressive samples.

## REFERENCES

- [1] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25(2), pp. 83-91, March 2008.
- [2] A. A. Moghadam, M. Aghagolzadeh, M. Kumar and H. Radha, "Compressive Demosaicing," *IEEE Multimedia Signal Processing (MMSP'10)*, 2010.
- [3] A. A. Moghadam, M. Aghagolzadeh, M. Kumar and H. Radha, "Compressive framework for demosaicing of natural images," *IEEE transactions on image processing*, vol. 22, issue 6, 2014.
- [4] A. A. Moghadam, M. Kumar and H. Radha, "Common and Innovative Visuals: A sparsity framework for video," *IEEE transactions on image processing*, 2014.
- [5] Aswin C. Sankaranarayanan, Pavan K. Turaga, Richard G. Baraniuk and Rama Chellappa, "Compressive Acquisition of Dynamic Scenes," *European Conference on Computer Vision*, Crete, Greece, September 2010.
- [6] E. Candes, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52(2) pp. 489-509, Feb. 2006.
- [7] David Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52(4), pp. 1289-1306, April 2006.
- [8] J. Mairal, M. Elad and G. Sapiro, "Sparse Representation for Color Image Restoration," *IEEE Transactions on Image Processing*, Vol. 17, No. 1, pages 53-69, January 2008.
- [9] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Preprint*, 2008.
- [10] J.A. Tropp and A. Gilbert, "Signal Recovery from Partial Information by Orthogonal Matching Pursuit," *IEEE Transactions on Information Theory*, Vol. 53, No. 12, pp. 4655-4666, Dec. 2007.
- [11] E. Candes and Y. Plan, "A probabilistic and RIPless theory of compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, issue 11.
- [12] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58 (1), pp. 267-288, 1996.
- [13] A. G. Hadigheh and T. Terlaky, "Sensitivity analysis in linear optimization: Invariant support set intervals," *European Journal of Operational Research*, vol. 169 (3), pp. 1158-1175.
- [14] A. G. Hadigheh, O. Romanko and T. Terlaky, "Sensitivity analysis in convex quadratic optimization: simultaneous perturbation of the objective and right-hand-side vectors," *Algorithmic Operations Research*, vol. 2 (2), pp. 94-111.
- [15] S. J. Wright, "Optimization Algorithms for Compressed Sensing," *SIAM Gator Student Conference*, March 2009.
- [16] S. Chandrasekaran and I. Ipsen, "Perturbation Theory for the Solution of Systems of Linear Equations," *Research Report, YALEU/DCS/RR-866*, October 1991.
- [17] A. Cohen, W. Dahmen and R. DeVore, "Compressed sensing and best k-term approximation," *Journal of American Mathematical Society*, vol. 22, pp. 211-231, 2009.
- [18] S.J. Kim, K. Koh, M. Lustig, S. Boyd and D. Gorinevsky, "An Interior-Point Method for Large-Scale  $\ell_1$ -Regularized Least Squares," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1(4), pp. 606-617, December 2007.
- [19] G. Tang and A. Nehorai, "Performance analysis for sparse support recovery," *IEEE transactions on information theory*, vol. 56, issue 3, 2010.
- [20] A. Karbasi, A. Hormati, S. Mohajer and Martin Vetterli, "Support recovery in compressed sensing: An estimation theoretic approach," *IEEE International Symposium on Information Theory (ISIT)*, 2009.
- [21] A. A. Moghadam and H. Radha, "Sparse Expander-like Real Valued Projection (SERP) matrices for compressed sensing," *IEEE globalSIP*, 2013.